REMARKS

Claims 1-12 are all the claims pending in the application. Applicants have added claim 13 by way of this amendment.

As preliminary matter, the Applicants asks the Examiner to acknowledge the claim of priority under 35 U.S.C. § 119 and to acknowledge receipt of certified copies of all of the priority documents. Further, the Examiner did not indicate whether the drawings are accepted or objected to. Applicants requests that the Examiner make this indication.

I. Status of the Application

Claims 1-12 have been examined. Claims 1-7 are rejected under 35 U.S.C. §112, second paragraph, and claims 1-12 are rejected under 35 U.S.C. §103(a).

II. The Present Invention

The invention has two subsystems a pre-processing subsystem and a tabulation subsystem. These subsystems are illustrated using the figures and tables provided in the specification.

The following is an illustration of the pre-processing subsystem, which enables a fast tabulation of data at a later time.

Table 1 illustrates a data file 20. *See Figure 1.* Each row corresponds to a record. Each column corresponds to a field. The column is the set of all instances of a given field from all records in a table. Data values are stored in the cells defined by a particular column and record.

By way of illustration, there are three fields and five records in this table. The third record in the AREA field has the data value "B" stored within. The SEX field has two distinct data values, which are "M" and "F." The AREA field has two distinct values, which are "A" and "B."

The data file 20 is provided to a field separator, which separates each field into an individual file 12. *See Fig. 1.* For example, Table 1 would be separated into three data files each with one column and five rows. Thereafter, the individual field files 12 are submitted to a distinct code mapping unit 14. The distinct code mapping unit reads through the individual field files and assigns incremental code for each data value found therein. This numerical code starts from 0 and increments by 1 with each distinct data value found. For Example, the encoding for the individual field files for the SEX and AREA field is as follows:

$SEX(K_1)$

M	Encode ⇒	0
M		0
F		1
F		1
M		0

$AREA(K_2)$

A	Encode ⇒	0
B		1
B		1
B		1
B		1

The encoded field files for SEX and AREA each have five records and two distinct values (i.e. 0 and 1). The number of distinct values of the SEX field is denoted D_1 and the number of distinct values of the AREA field is denoted D_2.

Further, a mapping file is created which stores the correspondence information. Mapping is defined as a rule of correspondence established between sets that associates each element of a set with an element in the same or another set. Correspondence is defined as a function such that for every element of one set there is a unique element of another set.

Applying the definitions to our example, there is a rule of correspondence established between the individual field file and the encoded field file. This rule of correspondence is stored in the mapping file 18. The mapping file 18 for the SEX field would store the following rule of correspondence:

$$f(SEX) = \begin{cases} 0 & \text{if } SEX = M \\ 1 & \text{if } SEX = F. \end{cases}$$

The tabulation subsystem is now illustrated. The tabulation subsystem uses the files created in the pre-processing subsystem to enable faster tabulation of data. For example, the tabulation subsystem receives requests for tabulation. A tabulation request specifies a plurality of field tuples (e.g. sex, area), where each field tuple specifies a field of the data file.

Upon receiving the request, the tabulation unit formulates an empty result array in the form of a one dimensional array for storing integer values. The length of the array in our

example would be 4, which is equal to $D_1 \times D_2$. Each cell in the array is used to store the number of records for one combination of the chosen field tuples. For example, the first cell may store the number of records with the combination M and A, which is easily seen from Table 1 to be one.

The cells of the result array are initialized to zero. The cells are then incremented according to the result of the computed index I, where the computed index I is determined based on the numerical identifiers (K_i where i : 1 to N) for the record in the selected fields and the number of distinct values (D_i where i : 1 to N-1) in the selected fields. *See Page 7, line 9.* Note that the possible values of I when there are four cells are 0, 1, 2, 3. Of course, if the result array had M cells than the possible values of I would be 0, ..., M-1, where each value of I would correspond to a distinct combination of the chosen field tuples. For our example:

K_1(SEX)	**K_2(AREA)**	**$I = K_1 + D_1 K_2$**	**I**
0	0	0 + 2 x 0	0
0	1	0 + 2 x 1	2
1	1	1 + 2 x 1	3
1	1	1 + 2 x 1	3
0	1	0 + 2 x 1	2

and the result array after incrementing the appropriate cells would be as follows:

I=0	I=1	I=2	I=3
1	0	2	2

It can be easily seen that cell 0 stores the number of distinct records with the combination M and A, cell 1 stores the number of distinct records of F and A, etc. Thus a sample tabulation as illustrated in Table 2 can easily be created from result array.

Although this example contains a small number of fields and records, it is apparent that for extremely large number of fields and records this method would enable faster data tabulation.

III. Indefiniteness Rejection Under 35 U.S.C. § 112, Second Paragraph

Claims 1-12 are rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which Applicants regard as the invention. In particular, the Examiner states that in claim 1 the phrase "correspondence" is indefinite and needs to be revised, and that it is not clear how "fields" relate to data and mapping.

Applicants submit that the term "correspondence" is definite, and further it is clear how the term "field" relates to data and mapping.

Mapping is defined as a rule of correspondence established between sets that associates each element of a set with an element in the same or another set. See *The American Heritage® Dictionary of the English Language, Fourth Edition Copyright © 2000 by Houghton Mifflin Company.* Correspondence is defined as a function such that for every element of one set there is a unique element of another set. *See WordNet 1.7.1 Copyright © 2001 by Princeton University.*

In the present invention, a mapping file is created which stores the correspondence information. Applying the definitions to our example above, there is a rule of correspondence established between the individual field file 12 and the encoded field file 16. This rule of correspondence is stored in the mapping file 18. For Example, the mapping file 18 for the SEX field would store the following rule of correspondence:

$$f(\text{SEX}) = \begin{cases} 0 & \text{if SEX} = \text{M} \\ 1 & \text{if SEX} = \text{F}. \end{cases}$$

Thus, the set of data values stored in the column that represents the SEX field are {M, M, F, F, M}. The set of data values stored in the column that represents the encoded SEX field are {0, 1}. The SEX field and the encoded SEX field are sets (domain and co-domain) defined by a function f from the SEX field set to the encoded SEX field set.

Applicants respectfully submit that the term field in relation to the data and the mapping files would be clear to a hypothetical person possessing the ordinary skill level in the pertinent art, data tabulation processing.

IV. Claim Rejections - 35 U.S.C. §103(a)

The Examiner has rejected claims 1-12 under 35 U.S.C. §103(a) as being unpatentable over Sampson et al. (US 5,212,639) in view of Rees et al. (US 5,748,878). Applicants respectfully traverse this rejection.

Please note that Applicants have amended claim 1 so that it is in its original form.

A. Claims 1 and 8

Applicants respectfully submit that claims 1 and 8 are patentable over the cited references, alone or in combination.

For example, claims 1 and 8 recite that for each record, a cell of a result array is determined based on the numerical identifiers for that record, and the result array cell incremented.

In the present invention, after each record is processed as recited in claims 1 and 8, the result array will contain the tabulation results. *See Specification: page 7, lines 11-20.*

The Examiner argues that Figure 5 of Sampson discloses that for each record, a cell of a result array is determined based on the numerical identifiers for that record, and the result array cell incremented.

Although, Figure 5 of Sampson discloses an array, the array is not a result array. The array disclosed is an array of numerical identifiers to be assigned to the account numbers in the chart of accounts, which is a preprocessing stage and not a tabulation stage. These numerical identifiers are later to be assigned to journal entries with corresponding account numbers. *See Col. 9, lines 21-35.* Further, the cells in the array of Figure 5 are not identified and incremented according to a numerical identifiers for that record, the cells of Fig. 5 are the numerical identifiers. The array disclosed in Sampson only teaches that each cell of an array may hold numerical identifiers corresponding to a chart of accounts.

Furthermore, the Examiner argues that Rees discloses a coding processor and a data reduction processor. *See Fig. 3.* Further, the Examiner argues that it would have been obvious to a person of ordinary skill in the art to have modified Sampson's system by implementing tabulation and coding processors for faster compilation of data.

Rees discloses tracking the execution of tags embedded in a software to perform a variety of analysis functions in essentially real time. *See Abstract.* The Examiner has not pointed to any disclosures in Rees that teach or suggest that for each record, a cell of a result array is determined based on the numerical identifiers for that record, and the result array cell

incremented. Applicants respectfully submit that Rees does not make up for the deficiencies of Sampson.

Neither reference discloses the feature claimed above, alone or in combination. Thus, claims 1 and 8 are patentable at least for this reason.

B. Claims 3 and 10

Applicants respectfully submit that claims 3 and 10 are patentable over the cited references, alone or in combination.

For example, claims 3 and 10 recite generating a plurality of encoded data files, one for each field.

The Examiner argues that Sampson discloses expanding the parental set by assigning numerical identifiers to the data entries from a data file containing the numerical identifiers for the data values in each field. *See Col. 3,lines 34-36.* However, Applicants submit that the Examiner is misinterpreting or misapplying the reference.

The present claims do not recite expanding the data file by assigning numerical identifiers to the data entries. In the present claims the data file is separated into individual files for each field, and then each individual file is encoded. Thus, the present claims generate an encoded data files for each field, and do not expand the data file by assigning numerical identifiers. *See Specification: page 5, lines 9-15.*

not in claims, not clear!

It seems that neither reference teaches or suggests a feature as claimed. Further, the Examiner has not pointed to any disclosure in the references that teaches or suggests such a feature.

Accordingly, claims 3 and 10 should be separately patentable.

C. Claim 5

Applicants respectfully submit that claim 5 is patentable over the cited references, alone or in combination.

For example, claim 5 recites initializing the result array to have a number of cells determined by the product of the number of numerical identifiers in the data fields.

The Examiner argues that Sampson discloses a result array. *See Fig. 5.* Further, the Examiner argues that it is common technique to initialize arrays to have a starting point and to maintain available memory for use as needed. However, Applicants respectfully submit that the Examiner is misinterpreting or misapplying the reference.

Although, Figure 5 of Sampson discloses an array, the array is not a result array. The array disclosed is an array of numerical identifiers to be assigned to the account numbers in the chart of accounts. These numerical identifiers are assigned to journal entries of the parental set with corresponding account numbers to form the expanded parental set.

In the present claims the number of cells in the array is determined by the product of the number of distinct values in each field specified in a tabulation request. *See Specification: page*

6, lines 20-26. The number of distinct values in each field being equal to the number of numerical identifiers in each field.

Neither reference teaches or suggests the initializing the number of cells as claimed. Further, the Examiner has not pointed to any disclosure in the references that teaches or suggests initializing the number of cells as claimed.

Accordingly, claim 5 should be separately patentable.

D. Claims 7 and 12

Applicants respectfully submit that claims 7 and 12 are patentable over the cited references, alone or in combination.

For example, claims 7 and 12 recite that a cell of the result array is identified for each data record by the value taken on by the cell identity I, where the cell identity is calculated using an equation defined by the numerical identifiers of the selected fields and the number of distinct values in the selected fields.

In the rejection of claims 7 and 12, the Examiner argues that Sampson discloses a result array. *See Fig. 5.* Further, the Examiner argues that it is common technique to initialize arrays to have a starting point and to maintain available memory for use as needed. However, Applicants submit that the Examiner is misinterpreting or misapplying the reference.

Although, Figure 5 of Sampson discloses an array, the array is not a result array. The array disclosed is an array of numerical identifiers to be assigned to the account numbers in the

chart of accounts. *See Col. 9, lines 21-35.* These numerical identifiers are assigned to journal entries with corresponding account numbers. Further, the cells in the array of Figure 5 are not identified and incremented according to a cell identity equation as in claims 7 and 12.

Neither reference teaches or suggests a method for identifying the cells of the result array by the value I, which is generated by applying the claimed equation to the specified fields in a record. Further, neither reference teaches or suggests that according to the value of the cell identity calculated by the claimed equation a corresponding cell is incremented.

Claims 7 and 12 are separately patentable over the references, since the references do not teach or suggest the features as claimed.

V. Newly added claims

Claim 13 has been added to provide more varied protection of the present invention. The claim is fully supported in the specification, and does not add new subject matter.

VI. Conclusion

In view of the above, reconsideration and allowance of this application are now believed to be in order, and such actions are hereby solicited. If any points remain in issue which the Examiner feels may be best resolved through a personal or telephone interview, the Examiner is kindly requested to contact the undersigned attorney at the telephone number listed below.

The USPTO is directed and authorized to charge all required fees, except for the Issue Fee and the Publication Fee, to Deposit Account No. 19-4880. Please also credit any overpayments to said Deposit Account.

Respectfully submitted,

Richard C. Turner
Registration No. 29,710

SUGHRUE MION, PLLC
Telephone: (202) 293-7060
Facsimile: (202) 293-7860

WASHINGTON OFFICE

23373
PATENT TRADEMARK OFFICE

Date: May 20, 2003

APPENDIX
VERSION WITH MARKINGS TO SHOW CHANGES MADE

IN THE CLAIMS:

The claims are amended as follows:

1. (Twice Amended) A method for data tabulation processing of a data file having a plurality of records in a plurality of data fields, comprising:

i) a pre-processing stage in which, for each individual data field, each distinct data value is identified and allocated a numerical identifier unique for that field[, wherein the preprocessing stage includes generating from said data file a mapping file which stores a correspondence between each of the distinct data values in the fields and the corresponding numerical identifiers]; and

ii) a tabulation stage in which, for each data record, a cell of a result array is determined based on the numerical identifiers for that record, and the result array cell incremented.

2. (Twice Amended) A method as in claim 1, wherein the preprocessing stage includes generating from said data file an encoded data file containing the numerical identifiers for the data values in each field, and a mapping file which stores a correspondence between each of the distinct data values in the fields and the corresponding numerical identifiers.

Claim 13 is added as new claim.